Pyruvate Carboxylase from

Corynebacterium glutamicum

STATEMENT OF GOVERNMENT RIGHTS IN THE INVENTION

Part of the work performed during development of this invention utilized U.S. Government funds. The U.S. Government has certain rights in this invention.

BACKGROUND OF THE INVENTION

Field of the Invention

The present invention relates to a *Corynebacterium glutamicum* pyruvate carboxylase protein and to polynucleotides encoding this protein.

Background Information

Pyruvate carboxylate is an important anaplerotic enzyme replenishing oxaloacetate consumed for biosynthesis during growth, or lysine and glutamic acid production in industrial fermentations.

The two-step reaction mechanism catalyzed by pyruvate carboxylase is shown below:

$$MgATP + HCO_3^- + \text{ENZ-biotin} \xrightarrow{Mg^{2+}\text{acetyl-CoA}} MgADP + Pi + \text{ENZ-biotin-}CO_2 \quad (1)$$

$$\text{ENZ-biotin-}CO_2 + \text{Pyruvate} \longrightarrow \text{ENZ-biotin} + \text{oxaloacetate} \quad (2)$$

In reaction (1) the ATP-dependent biotin carboxylase domain carboxylates a biotin prosthetic group linked to a specific lysine residue in the biotin-carboxyl-carrier protein (BCCP) domain. Acetyl-coenzyme A activates reaction (1) by increasing the rate

of bicarbonate-dependent ATP cleavage. In reaction (2), the BCCP domain donates the CO_2 to pyruvate in a reaction catalyzed by the transcarboxylase domain (Attwood, P.V., *Int. J. Biochem. Cell. Biol. 27*:231-249 (1995)).

Pruvate carboxylase genes have been cloned and sequenced from: *Rhizobium etli* (Dunn, M.F., *et al., J. Bacteriol. 178*:5960-5970 (1996)), *Bacillus stearothermophilus* (Kondo, H., *et al., Gene 191*:47-50 (1997), *Bacillus subtillis* (Genbank accession no. Z97025), *Mycobacterium tuberculosis* (Genbank accession no. Z83018), and *Methanobacterium thermoautotrophicum* (Mukhopadhyay, B., *J. Biol. Chem. 273*:5155-5166 (1998). Pyruvate carboxylase activity has been measured previously in *Brevibacterium lactofermentum* (Tosaka, O., *et al., Agric. Biol. Chem. 43*:1513-1519 (1979)) and *Corynebacterium glutamicum* (Peters-Wendisch, P.G., *et al., Microbiology 143*:1095-1103 (1997)).

Previous research has indicated that the yield and productivity of the aspartate family of amino acids depends critically on the carbon flux through anaplerotic pathways (Vallino, J.J., & Stephanopoulos, G., *Biotechnol. Bioeng. 41*:633-646 (1993)). On the basis of the metabolite balances, it can be shown that the rate of lysine production is less than or equal to the rate of oxaloacetate synthesis via the anaplerotic pathways.

SUMMARY OF THE INVENTION

The present invention provides isolated nucleic acid molecules comprising a polynucleotide encoding a pyruvate carboxylase polypeptide having the amino acid sequence in Figure 1 (SEQ ID NO:2) or the amino acid sequence encoded by the cosmid clone deposited in a bacterial host as ATCC Deposit Number _____. The nucleotide sequence determined by sequencing the deposited pyruvate carboxylase cosmid clone, which is shown in Figure 1 (SEQ ID NO:1), contains an open reading frame encoding a polypeptide of 1140 amino acid residues which has a deduced molecular weight of about 123.6 kDa. The 1140 amino acid sequence of the predicted pyruvate carboxylase protein is shown in Figure 1 and in SEQ ID NO:2.

Thus, one aspect of the invention provides an isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence selected from

the group consisting of: (a) a nucleotide sequence encoding the pyruvate carboxylase polypeptide having the complete amino acid sequence in SEQ ID NO:2; (b) a nucleotide sequence encoding the pyruvate carboxylase polypeptide having the complete amino acid sequence encoded by the cosmid clone contained in ATCC Deposit No. _____; and (c) a nucleotide sequence complementary to any of the nucleotide sequences in (a) or (b) above.

Further embodiments of the invention include isolated nucleic acid molecules that comprise a polynucleotide having a nucleotide sequence at least 90% identical, and more preferably at least 95%, 97%, 98% or 99% identical, to any of the nucleotide sequences in (a), (b) or (c) above, or a polynucleotide which hybridizes under stringent hybridization conditions to a polynucleotide having a nucleotide sequence identical to a nucleotide sequence in (a), (b) or (c), above. The polynucleotide which hybridizes does not hybridize under stringent hybridization conditions to a polynucleotide having a nucleotide sequence consisting of only A residues or of only T residues.

The present invention also relates to recombinant vectors which include the isolated nucleic acid molecules of the present invention and to host cells containing the recombinant vectors, as well as to methods of making such vectors and host cells and for using them for production of pyruvate carboxylase polypeptides or peptides by recombinant techniques.

The invention further provides an isolated pyruvate carboxylase polypeptide having amino acid sequence selected from the group consisting of: (a) the amino acid sequence of the pyruvate carboxylase polypeptide having the amino acid sequence shown in Figure 1 (SEQ ID NO:2); and (b) the amino acid sequence of the pyruvate carboxylase polypeptide having the complete amino acid sequence encoded by the cosmid clone contained in ATCC Deposit No. _____. The polypeptides of the present invention also include polypeptides having an amino acid sequence with at least 90% similarity, more preferably at least 95% similarity to those described in (a) or (b) above, as well as polypeptides having an amino acid sequence at least 70% identical, more preferably at least 90% identical, and still more preferably 95%, 97%, 98% or 99% identical to those above.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the nucleotide (SEQ ID NO:1) and deduced amino acid (SEQ ID NO:2) sequences of the complete pyruvate carboxylase protein determined by sequencing of the DNA clone contained in ATCC Deposit No. _____. The protein has sequence of about 1140 amino acid residues and a deduced molecular weight of about 123.6 kDa.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides isolated nucleic acid molecules comprising a polynucleotide encoding the pyruvate carboxylase protein having the amino acid sequence shown in Figure 1 (SEQ ID NO:2) which was determined by sequencing a cloned cosmid. The pyruvate carboxylase protein of the present invention shares sequence homology with *M. tuberculosis* and human pyruvate carboxylase proteins. The nucleotide sequence shown in Figure 1 (SEQ ID NO:1) was obtained by sequencing cosmid III F10 encoding a pyruvate carboxylase polypeptide, which was deposited on___ at the American Type Culture Collection, 10801 University Blvd., Manassas, VA 20110-2209, and given accession number _____.

Nucleic Acid Molecules

Unless otherwise indicated, all nucleotide sequences determined by sequencing a DNA molecule herein were determined using an automated DNA sequencer (such as the ABI Prism 377), and all amino acid sequences of polypeptides encoded by DNA molecules determined herein were predicted by translation of a DNA sequence determined as above. Therefore, as is known in the art for any DNA sequence determined by this automated approach, any nucleotide sequence determined herein may contain some errors. Nucleotide sequences determined by automation are typically at least about 90% identical, more typically at least about 95% to at least about 99.9% identical to the actual nucleotide

sequence of the sequenced DNA molecule. The actual sequence can be more precisely determined by other approaches including manual DNA sequencing methods well known in the art. As is also known in the art, a single insertion or deletion in a determined nucleotide sequence compared to the actual sequence will cause a frame shift in translation of the nucleotide sequence such that the predicted amino acid sequence encoded by a determined nucleotide sequence will be completely different from the amino acid sequence actually encoded by the sequenced DNA molecule, beginning at the point of such an insertion or deletion.

Unless otherwise indicated, each "nucleotide sequence" set forth herein is presented as a sequence of deoxyribonucleotides (abbreviated A, G , C and T). However, by "nucleotide sequence" of a nucleic acid molecule or polynucleotide is intended, for a DNA molecule or polynucleotide, a sequence of deoxyribonucleotides, and for an RNA molecule or polynucleotide, the corresponding sequence of ribonucleotides (A, G, C and U) where each thymidine deoxynucleotide (T) in the specified deoxynucleotide sequence in is replaced by the ribonucleotide uridine (U). For instance, reference to an RNA molecule having the sequence of SEQ ID NO:1 set forth using deoxyribonucleotide abbreviations is intended to indicate an RNA molecule having a sequence in which each deoxynucleotide A, G or C of SEQ ID NO:1 has been replaced by the corresponding ribonucleotide A, G or C, and each deoxynucleotide T has been replaced by a ribonucleotide U.

Using the information provided herein, such as the nucleotide sequence in Figure 1, a nucleic acid molecule of the present invention encoding a pyruvate carboxylase polypeptide may be obtained using standard cloning and screening procedures, such as those for cloning DNAs using mRNA as starting material. The pyruvate carboxylase protein shown in Figure 1 (SEQ ID NO:2) is about 63% identical to *M. tuberculosis* and 44% identical to human. As one of ordinary skill would appreciate, due to the possibilities of sequencing errors discussed above, as well as the variability of cleavage sites for leaders in different known proteins, the actual pyruvate carboxylase polypeptide encoded by the deposited cosmid comprises about 1140 amino acids, but may be anywhere in the range of 1133-1147 amino acids.

As indicated, nucleic acid molecules of the present invention may be in the form of RNA, such as mRNA, or in the form of DNA, including, for instance, cDNA and genomic DNA obtained by cloning or produced synthetically. The DNA may be double-stranded or single-stranded. Single-stranded DNA or RNA may be the coding strand, also known as the sense strand, or it may be the non-coding strand, also referred to as the anti-sense strand.

By "isolated" nucleic acid molecule(s) is intended a nucleic acid molecule, DNA or RNA, which has been removed from its native environment. For example, recombinant DNA molecules contained in a vector are considered isolated for the purposes of the present invention. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of the DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

Isolated nucleic acid molecules of the present invention include DNA molecules comprising an open reading frame (ORF) with an initiation codon at positions 199-201 of the nucleotide sequence shown in Figure 1 (SEQ ID NO:1); DNA molecules comprising the coding sequence for the pyruvate carboxylase protein shown in Figure 1 and SEQ ID NO:2; and DNA molecules which comprise a sequence substantially different from those described above but which, due to the degeneracy of the genetic code, still encode the pyruvate carboxylase protein. Of course, the genetic code is well known in the art. Thus, it would be routine for one skilled in the art to generate the degenerate variants described above.

In another aspect, the invention provides isolated nucleic acid molecules encoding the pyruvate carboxylase polypeptide having an amino acid sequence encoded by the cosmid clone deposited as ATCC Deposit No. _____. Preferably, this nucleic acid molecule will encode the polypeptide encoded by the above-described deposited clone. The invention further provides an isolated nucleic acid molecule having the nucleotide sequence shown in Figure 1 (SEQ ID NO:1) or the nucleotide sequence of the pyruvate carboxylase DNA contained in the above-described deposited clone, or nucleic acid molecule having a sequence

complementary to one of the above sequences.

In another aspect, the invention provides an isolated nucleic acid molecule comprising a polynucleotide which hybridizes under stringent hybridization conditions to a portion of the polynucleotide in a nucleic acid molecule of the invention described above, for instance, the cosmid clone contained in ATCC Deposit _____. By "stringent hybridization conditions" is intended overnight incubation at 42°C in a solution comprising: 50% formamide, 5x SSC (150 mM NaCl, 15mM trisodium citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 μg/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1x SSC at about 65°C. By a polynucleotide which hybridizes to a "portion" of a polynucleotide is intended a polynucleotide (either DNA or RNA) hybridizing to at least about 15 nucleotides (nt), and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably about 30-70 nt of the reference polynucleotide. These are useful as diagnostic probes and primers.

Of course, polynucleotides hybridizing to a larger portion of the reference polynucleotide (e.g., the deposited cosmid clone), for instance, a portion 50-750 nt in length, or even to the entire length of the reference polynucleotide, also useful as probes according to the present invention, as are polynucleotides corresponding to most, if not all, of the nucleotide sequence of the deposited DNA or the nucleotide sequence as shown in Figure 1 (SEQ ID NO:1). By a portion of a polynucleotide of "at least 20 nt in length," for example, is intended 20 or more contiguous nucleotides from the nucleotide sequence of the reference polynucleotide, (e.g., the deposited DNA or the nucleotide sequence as shown in Figure 1 (SEQ ID NO:1)). As indicated, such portions are useful diagnostically either as a probe according to conventional DNA hybridization techniques or as primers for amplification of a target sequence by the polymerase chain reaction (PCR), as described, for instance, in *Molecular Cloning, A Laboratory Manual*, 2nd. edition, edited by Sambrook, J., Fritsch, E. F. and Maniatis, T., (1989), Cold Spring Harbor Laboratory Press, the entire disclosure of which is hereby incorporated herein by reference.

Since a pyruvate carboxylase cosmid clone has been deposited and its

determined nucleotide sequence is provided in Figure 1 (SEQ ID NO:1), generating polynucleotides which hybridize to a portion of the pyruvate carboxylase DNA molecule would be routine to the skilled artisan. For example, restriction endonuclease cleavage or shearing by sonication of the pyruvate carboxylase cosmid clone could easily be used to generate DNA portions of various sizes which are polynucleotides that hybridize to a portion of the pyruvate carboxylase DNA molecule. Alternatively, the hybridizing polynucleotides of the present invention could be generated synthetically according to known techniques.

As indicated, nucleic acid molecules of the present invention which encode the pyruvate carboxylase protein polypeptide may include, but are not limited to those encoding the amino acid sequence of the polypeptide, by itself; the coding sequence for the polypeptide and additional sequences, such as a pre-, or pro- or prepro- protein sequence; the coding sequence of the polypeptide, with or without the aforementioned additional coding sequences, together with additional, non-coding sequences, including for example, but not limited to introns and non-coding 5' and 3' sequences, such as the transcribed, non-translated sequences that play a role in transcription, mRNA processing - including splicing and polyadenylation signals, for example - ribosome binding and stability of mRNA; an additional coding sequence which codes for additional amino acids, such as those which provide additional functionalities. Thus, the sequence encoding the polypeptide may be fused to a marker sequence, such as a sequence encoding a peptide which facilitates purification of the fused polypeptide. In certain preferred embodiments of this aspect of the invention, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (Qiagen, Inc.), among others, many of which are commercially available. As described in Gentz *et al.*, *Proc. Natl. Acad. Sci., USA 86*:821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein. The "HA" tag is another peptide useful for purification which corresponds to an epitope derived from the influenza hemagglutinin protein, which has been described by Wilson *et al.*, *Cell 37*: 767 (1984).

The present invention further relates to variants of the nucleic acid molecules of the present invention, which encode portions, analogs or derivatives

of the pyruvate carboxylase protein. Variants may occur naturally, such as a natural allelic variant. By an "allelic variant" is intended one of several alternate forms of a gene occupying a given locus on a chromosome of an organism. *Genes II*, Lewin, ed. Non-naturally occurring variants may be produced using art-known mutagenesis techniques.

Such variants include those produced by nucleotide substitutions, deletions or additions. The substitutions, deletions or additions may involve one or more nucleotides. The variants may be altered in coding or non-coding regions or both. Alterations in the coding regions may produce conservative or non-conservative amino acid substitutions, deletions or additions. Especially preferred among these are silent substitutions, additions and deletions, which do not alter the properties and activities of the pyruvate carboxylase protein or portions thereof. Also especially preferred in this regard are conservative substitutions. Most highly preferred are nucleic acid molecules encoding the pyruvate carboxylase protein having the amino acid sequence shown in Figure 1 (SEQ ID NO:2).

Also preferred are mutants or variants whereby preferably pyruvate carboxylase is expressed 2 to 20 fold higher than its expression in *C. glutamicum* as well as feedback inhibition mutants.

Further embodiments of the invention include isolated nucleic acid molecules comprising a polynucleotide having a nucleotide sequence at least 90% identical, and more preferably at least 95%, 97%, 98% or 99% identical to (a) a nucleotide sequence encoding the pyruvate carboxylase polypeptide having the complete amino acid sequence in SEQ ID NO:2; (b) a nucleotide sequence encoding the pyruvate carboxylase polypeptide having the complete amino acid sequence encoded by the cosmid clone contained in ATCC Deposit No. _____; or (c) a nucleotide sequence complementary to any of the nucleotide sequences in (a) or (b).

By a polynucleotide having a nucleotide sequence at least, for example, 95% "identical" to a reference nucleotide sequence encoding a pyruvate carboxylase polypeptide is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference

nucleotide sequence encoding the pyruvate carboxylase polypeptide. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. These mutations of the reference sequence may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence.

As a practical matter, whether any particular nucleic acid molecule is at least 90%, 95%, 97%, 98% or 99% identical to, for instance, the nucleotide sequence shown in Figure 1 or to the nucleotides sequence of the deposited cosmid clone can be determined conventionally using known computer programs such as the Bestfit program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711). Bestfit uses the local homology algorithm of Smith and Waterman (*Advances in Applied Mathematics 2*: 482-489 (1981)) to find the best segment of homology between two sequences. When using Bestfit or any other sequence alignment program to determine whether a particular sequence is, for instance, 95% identical to a reference sequence according to the present invention, the parameters are set, of course, such that the percentage of identity is calculated over the full length of the reference nucleotide sequence and that gaps in homology of up to 5% of the total number of nucleotides in the reference sequence are allowed.

The present application is directed to nucleic acid molecules at least 90%, 95%, 97%, 98% or 99% identical to the nucleic acid sequence shown in Figure 1 (SEQ ID NO:1) or to the nucleic acid sequence of the deposited DNA, irrespective of whether they encode a polypeptide having pyruvate carboxylase activity. This is because, even where a particular nucleic acid molecule does not encode a polypeptide having pyruvate carboxylase activity, one of skill in the art would still know how to use the nucleic acid molecule, for instance, as a hybridization probe or a polymerase chain reaction (PCR) primer.

Preferred, however, are nucleic acid molecules having sequences at least 90%, 95%, 97%, 98% or 99% identical to the nucleic acid sequence shown in Figure 1 (SEQ ID NO:1) or to the nucleic acid sequence of the deposited DNA which do, in fact, encode a polypeptide having pyruvate carboxylase protein activity. By "a polypeptide having pyruvate carboxylase activity" is intended polypeptides exhibiting activity similar, but not necessarily identical, to an activity of the pyruvate carboxylase protein of the invention as measured in a particular biological assay.

Of course, due to the degeneracy of the genetic code, one of ordinary skill in the art will immediately recognize that a large number of the nucleic acid molecules having a sequence at least 90%, 95%, 97%, 98%, or 99% identical to the nucleic acid sequence of the deposited DNA or the nucleic acid sequence shown in Figure 1 (SEQ ID NO:1) will encode a polypeptide "having pyruvate carboxylase protein activity." In fact, since degenerate variants of these nucleotide sequences all encode the same polypeptide, this will be clear to the skilled artisan even without performing the above described comparison assay. It will be further recognized in the art that, for such nucleic acid molecules that are not degenerate variants, a reasonable number will also encode a polypeptide having pyruvate carboxylase protein activity. This is because the skilled artisan is fully aware of amino acid substitutions that are either less likely or not likely to significantly effect protein function (e.g., replacing one aliphatic amino acid with a second aliphatic amino acid).

For example, guidance concerning how to make phenotypically silent amino acid substitutions is provided in Bowie, J. U., *et al.*, "Deciphering the Message in Protein Sequences: Tolerance to Amino Acid Substitutions," *Science 247:*1306-1310 (1990), wherein the authors indicate that there are two main approaches for studying the tolerance of an amino acid sequence to change. The first method relies on the process of evolution, in which mutations are either accepted or rejected by natural selection. The second approach uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene and selections or screens to identify sequences that maintain functionality. As the authors state, these studies have revealed that proteins are surprisingly tolerant of

amino acid substitutions. The authors further indicate which amino acid changes are likely to be permissive at a certain position of the protein. For example, most buried amino acid residues require nonpolar side chains, whereas few features of surface side chains are generally conserved. Other such phenotypically silent substitutions are described in Bowie, J.U., *et al., supra,* and the references cited therein.

Vectors and Host Cells

The present invention also relates to vectors which include the isolated DNA molecules of the present invention, host cells which are genetically engineered with the recombinant vectors, and the production of pyruvate carboxylase polypeptides or portions thereof by recombinant techniques.

Recombinant constructs may be introduced into host cells using well known techniques such as infection, transduction, transfection, transvection, conjugation, electroporation and transformation. The vector may be, for example, a phage, plasmid, viral or retroviral vector.

The polynucleotides may be joined to a vector containing a selectable marker for propagation in a host. Generally, a plasmid vector is introduced in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. If the vector is a virus, it may be packaged *in vitro* using an appropriate packaging cell line and then transduced into host cells.

Preferred are vectors comprising cis-acting control regions to the polynucleotide of interest. Appropriate trans-acting factors may be supplied by the host, supplied by a complementing vector or supplied by the vector itself upon introduction into the host.

In certain preferred embodiments in this regard, the vectors provide for specific expression, which may be inducible and/or cell type-specific. Particularly preferred among such vectors are those inducible by environmental factors that are easy to manipulate, such as temperature and nutrient additives.

Expression vectors useful in the present invention include chromosomal-, episomal- and virus-derived vectors, e.g., vectors derived from bacterial plasmids, bacteriophage, yeast episomes, yeast chromosomal elements, viruses such as

baculoviruses, papova viruses, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as cosmids and phagemids.

The DNA insert should be operatively linked to an appropriate promoter, such as the phage lambda P_L promoter, the *E. coli lac*, *trp* and *tac* promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled artisan. The expression constructs will further contain sites for transcription initiation, termination and, in the transcribed region, a ribosome binding site for translation. The coding portion of the mature transcripts expressed by the constructs will include a translation initiating codon (AUG or GUG) at the beginning and a termination codon appropriately positioned at the end of the polypeptide to be translated.

As indicated, the expression vectors will preferably include at least one selectable marker. Such markers include dihydrofolate reductase or neomycin resistance for eukaryotic cell culture and tetracycline, ampicillin, chloramphenicol or kanamycin resistance genes for culturing in *E. coli* and other bacteria. Representative examples of appropriate hosts include bacterial cells, such as *E. coli*, *C. glutamicum*, Streptomyces and *Salmonella typhimurium* cells; fungal cells, such as yeast cells. Appropriate culture media and conditions for the above-described host cells are known in the art.

Among vectors preferred for use in bacteria include pA2, pQE70, pQE60 and pQE-9, available from Qiagen; pBS vectors, Phagescript vectors, Bluescript vectors, pNH8A, pNH16a, pNH18A, pNH46A, available from Stratagene; and ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia. Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1 and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. Other suitable vectors will be readily apparent to the skilled artisan.

Among known bacterial promoters suitable for use in the present invention include the *E. coli lac*I and *lac*Z promoters, the T3 and T7 promoters, the *gpt* promoter, the lambda P_R and P_L promoters and the *trp* promoter. Suitable eukaryotic promoters include the CMV immediate early promoter, the HSV

thymidine kinase promoter, the early and late SV40 promoters, the promoters of retroviral LTRs, such as those of the Rous sarcoma virus (RSV), and metallothionein promoters, such as the mouse metallothionein-I promoter.

Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection or other methods. Such methods are described in many standard laboratory manuals, such as Davis *et al.*, "Basic Methods in Molecular Biology," (1986).

Transcription of the DNA encoding the polypeptides of the present invention by higher eukaryotes may be increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp that act to increase transcriptional activity of a promoter in a given host cell-type. Examples of enhancers include the SV40 enhancer, which is located on the late side of the replication origin at bp 100 to 270, the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

For secretion of the translated protein into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment, appropriate secretion signals may be incorporated into the expressed polypeptide. The signals may be endogenous to the polypeptide or they may be heterologous signals.

The polypeptide may be expressed in a modified form, such as a fusion protein, and may include not only secretion signals, but also additional heterologous functional regions. Thus, for instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence in the host cell, during purification, or during subsequent handling and storage. Also, peptide moieties may be added to the polypeptide to facilitate purification.

The pyruvate carboxylase protein can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography,

affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification.

Polypeptides of the present invention include naturally purified products, products of chemical synthetic procedures, and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes.

Pyruvate Carboxylase Polypeptides and Peptides

The invention further provides an isolated pyruvate carboxylase polypeptide having the amino acid sequence encoded by the deposited DNA, or the amino acid sequence in Figure 1 (SEQ ID NO:2), or a peptide or polypeptide comprising a portion of the above polypeptides. The terms "peptide" and "oligopeptide" are considered synonymous (as is commonly recognized) and each term can be used interchangeably as the context requires to indicate a chain of at least to amino acids coupled by peptidyl linkages. The word "polypeptide" is used herein for chains containing more than ten amino acid residues. All oligopeptide and polypeptide formulas or sequences herein are written from left to right and in the direction from amino terminus to carboxy terminus.

It will be recognized in the art that some amino acid sequence of the pyruvate carboxylase polypeptide can be varied without significant effect on the structure or function of the protein. If such differences in sequence are contemplated, it should be remembered that there will be critical areas on the protein which determine activity. In general, it is possible to replace residues which form the tertiary structure, provided that residues performing a similar function are used. In other instances, the type of residue may be completely unimportant if the alteration occurs at a non-critical region of the protein.

Thus, the invention further includes variations of the pyruvate carboxylase polypeptide which show substantial activity or which include regions of pyruvate carboxylase protein such as the protein portions discussed below. Such mutants include deletions, insertions, inversions, repeats, and type substitutions (for example, substituting one hydrophilic residue for another, but not strongly hydrophilic for strongly hydrophobic as a rule). Small changes or such "neutral" amino acid substitutions will generally have little effect on activity.

Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe, Tyr.

As indicated in detail above, further guidance concerning which amino acid changes are likely to be phenotypically silent (i.e., are not likely to have a significant deleterious effect on a function) can be found in Bowie, J.U., *et al.*, "Deciphering the Message in Protein Sequences: Tolerance to Amino Acid Substitutions," *Science 247:*1306-1310 (1990).

The polypeptides of the present invention are preferably provided in an isolated form, and preferably are substantially purified. A recombinantly produced version of the pyruvate carboxylase polypeptide can be substantially purified by the one-step method described in Smith and Johnson, *Gene 67:*31-40 (1988).

The polypeptides of the present invention include the polypeptide encoded by the deposited DNA, the polypeptide of SEQ ID NO:2, as well as polypeptides which have at least 90% similarity, more preferably at least 95% similarity, and still more preferably at least 97%, 98% or 99% similarity to those described above. Further polypeptides of the present invention include polypeptides at least 70% identical, more preferably at least 90% or 95% identical, still more preferably at least 97%, 98% or 99% identical to the polypeptide encoded by the deposited DNA, to the polypeptide of SEQ ID NO:2, and also include portions of such polypeptides with at least 30 amino acids and more preferably at least 50 amino acids.

By "% similarity" for two polypeptides is intended a similarity score

produced by comparing the amino acid sequences of the two polypeptides using the Bestfit program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711) and the default settings for determining similarity. Bestfit uses the local homology algorithm of Smith and Waterman (Advances in Applied Mathematics *2*: 482-489, 1981) to find the best segment of similarity between two sequences.

By a polypeptide having an amino acid sequence at least, for example, 95% "identical" to a reference amino acid sequence of a pyruvate carboxylase polypeptide is intended that the amino acid sequence of the polypeptide is identical to the reference sequence except that the polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the reference amino acid of the pyruvate carboxylase polypeptide. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a reference amino acid sequence, up to 5% of the amino acid residues in the reference sequence may be deleted or substituted with another amino acid, or a number of amino acids up to 5% of the total amino acid residues in the reference sequence may be inserted into the reference sequence. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

As a practical matter, whether any particular polypeptide is at least 90%, 95%, 97%, 98% or 99% identical to, for instance, the amino acid sequence shown in Figure 1 (SEQ ID NO:2) or to the amino acid sequence encoded by deposited cosmid clone can be determined conventionally using known computer programs such the Bestfit program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711. When using Bestfit or any other sequence alignment program to determine whether a particular sequence is, for instance, 95% identical to a reference sequence according to the present invention, the parameters are set, of course, such that the percentage of identity is calculated over the full length of the reference amino acid sequence and that gaps in homology of up to 5% of the total

number of amino acid residues in the reference sequence are allowed.

Genetic Tools for Manipulating Corynebacterium

To make the genetic changes necessary for metabolic engineering in *Corynebacterium,* researchers need to be able to identify and clone the genes that are involved in the target pathway. They also need methods for altering these genes to affect the regulation or level of expression of the enzymes they encode, and for subsequently reintroducing the altered genes into *Corynebacterium* to monitor their effects on amino acid biosynthesis. Therefore, metabolic engineers must have at their disposal an array of plasmids that can replicate in both *Corynebacterium* and other, more easily manipulated hosts, such as *E. coli*. Also required are a collection of selectable markers encoding, for example, antibiotic resistance, well-characterized transcriptional promoters that permit regulation of the altered genes, and efficient transformation or conjugation systems that allow the plasmids to be inserted into the target *Corynebacterium* strain.

Plasmids. Several different plasmids have been isolated and developed for the introduction and expression of genes in *Corynebacterium* (Sonnen, H., *et al.*, *Gene 107*:69-74 (1991)). The majority of these were originally identified as small (3-5 kbp), cryptic plasmids from *C. glutamicum*, *C. callunae*, and *C. lactofermentum*. They fall into four compatibility groups, exemplified by the plasmids pCC1, pBL1, pHM1519, and pGA1. Shuttle vectors, plasmids that are capable of replicating in both *Corynebacterium* and *E. coli,* have been developed from these cryptic plasmids by incorporating elements from known *E. coli* plasmids (particularly the ColE1 origin of replication from pBR322 or pUC18), as well as antibiotic-resistance markers. A fifth class of plasmids that is very useful for manipulating *Corynebacterium* is based on pNG2, a plasmid originally isolated from *Corynebacterium diphtheriae* (Serwold-Davis, T.M., *et al.*, *Proc. Natl. Acad. Sci. USA 84*:4964-4968 (1987)). This plasmid and its derivatives replicate efficiently in many species of corynebacteria, as well as in *E. coli*. Since the sole origin of replication in pNG2 (an element of only 1.8 kbp) functions in both the Gram-positive and Gram-negative host, there is no need to add an additional ColE1-type element to it. As a result, pNG2 derivatives (e.g., pEP2) are much smaller than other *Corynebacterium* shuttle

vectors and are therefore more easily manipulated.

Selectable Markers. Several genes conferring antibiotic resistance have proven useful for plasmid selection and in other recombinant DNA work in corynebacteria. These include the kanamycin resistance determinant from Tn*903*, a hygromycin resistance marker isolated from *Streptomyces hygroscopicus*, a tetracycline resistance gene from *Streptococcus faecalis*, a bleomycin resistance gene from Tn*5*, and a chloramphenicol resistance marker from *Streptomyces acrimycini*. The β-lactamase gene that is employed in many *E. coli* plasmids such as pBR322 does not confer ampicillin resistance in *Corynebacterium*.

Transformation Systems. Several methods have been devised for introducing foreign DNA into *Corynebacterium*. The earliest method to be employed routinely was based on protocols that had been successful for other Gram-positive species involving incubation of spheroplasts in the presence of DNA and polyethylene glycol (Yoshihama, M., *et al.*, *J. Bacteriol. 162*:591-597 (1985)). While useful, these methods were generally inefficient, often yielding fewer than 10^5 transformants per milligram of DNA. Electroporation of *Corynebacterium* spheroplasts has proven to be a much more efficient and reliable means of transformation. Spheroplasts are generated by growing the cells in rich media containing glycine and/or low concentrations of other inhibitors of cell wall biosynthesis, such as isonicotinic acid hydrazide (isoniazid), ampicillin, penicillin G, or Tween-80. The spheroplasts are then washed in low-salt buffers containing glycerol, concentrated, and mixed with DNA before being subjected to electroporation. Efficiencies as high as 10^7 transformants per microgram of plasmid DNA have been reported with this protocol.

A third method for DNA transfer into corynebacteria involves transconjugation. This method takes advantage of the promiscuity of *E. coli* strains carrying derivatives of the plasmid RP4. In *E. coli*, RP4 encodes many functions that mediate the conjugal transfer of plasmids from the host strain to other recipient strains of *E. coli*, or even to other species. These "*tra* functions" mediate pilus formation and plasmid transfer. RP4 also carries an origin of transfer, *oriT*, a *cis*-acting element that is recognized by the transfer apparatus that allows the plasmid to be conducted through the

pilus and into the recipient strain. From this system Simon *et al.* (*Bio/Technology 1*:784-791 (1985)) have developed a useful transconjugation tool that allows the transfer of plasmids from *E. coli* to *Corynebacterium*. They relocated the *tra* functions from RP4 into the *E. coli* chromosome in a strain called S17-1. Plasmids carrying the RP4 *oriT* can be mobilized from S17-1 into other recipients very efficiently. Although this method has proven useful for introducing replicating plasmids into *Corynebacterium*, it has proven even more useful for generating gene disruptions. This is accomplished by introducing a selectable marker into a clone of the *Corynebacterium* gene that is targeted for disruption. This construct is then ligated into an *E. coli* plasmid that carries the RP4 *oriT* but lacks an origin to support replication in *Corynebacterium*. S17-1 carrying this plasmid is then incubated with the recipient strain and the mixture is later transferred to a selective medium. Because the plasmid that was introduced is unable to replicate in corynebacteria, transconjugants that express the selectable marker are most likely to have undergone a cross-over recombination within the genomic DNA.

Restriction-Deficient Strains. Regardless of the transformation system used, there is clear precedent in the literature that corynebacteria are able to recognize *E. coli*-derived DNA as foreign and will most often degrade it. This ability has been attributed to the *Corynebacterium* restriction and modification system. To overcome this system, some transformation and transconjugation protocols call for briefly heating the recipient strain prior to transformation. The heat treatment presumably inactivates the enzymes responsible for the restriction system, allowing the introduced DNA to become established before the enzymes are turned over. Another strategy for improving the efficiency of DNA transfer has been to isolate *Corynebacterium* mutants that are deficient in the restriction system. These strains will incorporate plasmids that had been propagated in *E. coli* with almost the same efficiency as plasmids that had been propagated in *Corynebacterium*. In an alternate strategy used to circumvent the restriction system in *Corynebacterium*, Leblon and coworkers (Reyes, O., *et al.*, *Gene 107*:61-68 (1991)) developed an "integron" system for gene disruption. Integrons are DNA molecules that have the same restriction/modification properties as the target host's DNA, carry DNA that is homologous to a portion of the host genome (i.e., a region of the genome that is to be disrupted), and are unable to replicate in the host cell. A cloned gene

from *Corynebacterium* is first interrupted with a selectable marker in a plasmid that is propagated in one *Cornynebacterium* strain. This construct is then excised from the corynebacterial plasmid and self-ligated to form a non-replicating circular molecule. This "integron" is then electroporated into the restrictive host. Modification of the DNA allows the integron to elude the host restriction system, and recombination into the host genome permits expression of the selectable marker.

Promoters. Reliable transcriptional promoters are required for efficient expression of foreign genes in *Corynebacterium*. For certain experiments, there is also a need for regulated promoters whose activity can be induced under specific culture conditions. Promoters such as the *fda*, *thrC*, and *hom* promoters derived from *Corynebacterium* genes have proven useful for heterologous gene expression. Inducible promoters from *E. coli*, such as P_{lac}, and P_{trc}, which are induced by isopropylthiogalactopyranoside (IPTG) when the *lac* repressor (*lacI*) is present; P_{trp}, which responds to the inducer indole acrylic acid when the *trp* repressor (*trpR*) is present; and lambda P_L, which is repressed in the presence of the temperature-sensitive lambda repressor (cI857), have all been used to modulate gene expression in *Corynebacterium*.

Gene Identification. With all other genetic tools in place, there still remains the challenge of identifying relevant genes from *Corynebacterium*. In *E. coli*, some of the resources that have been used to isolate genes are transducing phage, transposable elements, genetic maps of the *E. coli* chromosome from transduction and transconjugation experiments, and more recently, complete physical and sequence maps of the chromosome. To date, the most successful method for identifying and recovering genes from *Corynebacterium* has been to use *Corynebacterium* genomic DNA to complement known auxotrophs of *E. coli*. In this exercise, libraries of plasmids carrying fragments of the *Corynebacterium* genome are introduced into *E. coli* strains that are deficient in a particular enzyme or function. Transformants that no longer display the auxotrophy (e.g., homoserine deficiency) are likely to carry the complementing gene from *Corynebacterium*. This strategy has led to isolation of numerous *Corynebacterium* genes, including several from the pathways responsible for synthesis of aspartate-derived and aromatic amino acids, intermediary metabolism, and other cellular processes. One

limitation to this strategy is that not all genes from *Corynebacterium* will be expressed in the *E. coli* host. Thus, although a gene may be represented in the plasmid library, it may be unable to complement the *E. coli* mutation and therefore would not be recovered during selection. Overcoming this limitation, a smaller number of genes have been identified with a similar strategy in which a plasmid library from wild-type *Corynebacterium* was used to directly complement mutations in other *Corynebacterium* strains. Although this strategy avoids the concern of insufficient gene expression in the auxotrophic host, its utility is limited by poor plasmid-transformation efficiency in the auxotrophs. Still other genes have been identified by hybridization with nucleic acid probes based upon homologous genes from other species, and direct amplification of genes using the polymerase chain reaction and degenerate oligonucleotide primers.

Transposable Elements. Transposable elements are extremely powerful tools in gene identification because they couple mutagenesis with gene recovery. Unlike classical mutagenesis techniques, which generate point mutations or small deletions within a gene, when transposable elements insert within a gene they form large disruptions, thereby "tagging" the altered gene for easier identification. A number of transposable elements have been found to transpose in *Corynebacterium*. Transposons found in the plasmids pTP10 of *C. xerosis* and pNG2 of *C. diphtheriae* have been shown to transpose in *C. glutamicum* and confer resistance to erythromycin. A group from the Mitsubishi Chemical Company in Japan developed a series of artificial transposons from an insertion sequence, IS31831, that they discovered in *C. glutamicum* (Vertes, A.A., *et al.*, *Mol. Gen. Genet. 245*:397-405 (1994)). After inserting a selectable marker between the inverted repeats of IS31831, these researchers were able to introduce the resulting transposon into *C. glutamicum* strains on an *E. coli* plasmid (unable to replicate in *Corynebacterium)* via electroporation. They found that the selectable marker had inserted into the genome of the target cell at a frequency of approximately 4×10^4 mutants/μg DNA. The use of such transposons to generate *Corynebacterium* auxotrophs has led to the isolation of several genes responsible for amino acid biosynthesis, as well as other functions in corynebacteria.

Transducing Phage. Transducing phage have been used in other systems for mapping

genetic loci and for isolating genes. In 1976, researchers at Ajinomoto Co. in Japan surveyed 150 strains of characterized and uncharacterized strains of glutamic acid-producing coryneform bacteria to identify phage that might be useful for transduction (Mornose, H., *et al.*, *J. Gen. Appl. Microbiol. Rev. 16*:243-252 (1995)). Of 24 different phage isolates recovered from this screen, only three were able to transduce a *trp* marker from a trp^+ donor to a trp^- recipient with any appreciable frequency, although even this efficiency was only 10^{-7} or less. These researchers were able to improve transduction efficiency slightly by including 4 mM cyclic adenosine monophosphate (cAMP) or 1.2 M magnesium chloride. Several different researchers have attempted to develop reliable transduction methods by isolating corynephages from sources such as contaminated industrial fermentations, soil, and animal waste. Although many phage have been isolated and characterized, few have been associated with transduction, and an opportunity still exists to develop a reliable, high-efficiency transduction system for general use with the glutamic acid-producing bacteria.

EXAMPLES

The following protocols and experimental details are referenced in the examples that follow.

Bacterial strains and plasmids

C. glutamicum 21253 (*hom*$^-$, lysine overproducer) was used for the preparation of chromosomal DNA. *Escherichia coli* DH5α (*hsdR*$^-$, *recA*$^-$) (Hanahan, D., *J. Mol. Biol. 166*:557-580 (1983)) was used for transformations. Plasmid pCR2.1 TOPO (Invitrogen) was used for cloning polymerase chain reaction (PCR) products. The plasmid pRR850 was constructed in this study and contained an 850-bp PCR fragment cloned in the pCR2.1 TOPO plasmid.

Media and culture conditions

E. coli strains were grown in Luria-Bertani (LB) medium at 37°C (Sambrook, J., *et al., Molecular cloning: a laboratory manual*, 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1989)). *C. glutamicum* was grown in LB medium at 30°C. Where noted, ampicillin was used at the following concentrations: 100 µg/ml in plates and 50 µg/ml in liquid culture.

DNA manipulations

Genomic DNA was isolated from *C. glutamicum* as described by Tomioka *et al.* (Tomioka, N., *et al., Mol. Gen. Genet. 184*:359-363 (1981)). PCR fragments were cloned into the pCR2.1 TOPO vector following the manufacturer's instructions. Cosmid and plasmid DNA were prepared using Qiaprep spin columns and DNA was extracted from agarose gels with the Qiaex kit (Qiagen). For large-scale high-purity preparation of cosmid DNA for sequencing, the Promega Wizard kit was used (Promega). Standard techniques were used for transformation of *E. coli* and agarose gel electrophoresis (Sambrook, J., *et al., Molecular cloning: a laboratory manual*, 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1989)). Restriction enzymes were purchased from Boehringer Mannheim or New England Biolabs.

Cosmid library

The cosmid library used was constructed by cloning *C. glutamicum* chromosomal DNA into the Supercos vector (Stratagene).

Polymerase Chain Reaction (PCR)

PCR was performed using the Boehringer Mannheim PCR core kit following the manufacturer's instructions. When PCR was performed on *Corynebacterium* chromosomal DNA, about 1 µg DNA was used in each reaction. The forward primer used was

5'GTCTTCATCGAGATGAATCCGCG3' and the reverse primer used was

5'CGCAGCGCCACATCGTAAGTCGC3' for the PCR reaction.

Dot-blot analysis

Dot blots containing DNA from cosmids identified in this study and the probe as a positive control were prepared using the S&S (Schleicher & Schüll) minifold apparatus. An 850-bp fragment encoding a portion of the *C. glutamicum* pyruvate carboxylate gene was used as the probe. The probe was labeled with digoxigenin-11-dUTP (Boehringer Mannheim) in a randomly primed DNA-labeling reaction as described by the manufacturer. Hybridization, washing and colorimetric detection of the dot blots were done with the Genius system from Boehringer following the protocols in their user's guide for filter hybridization. The initial hybridization with the 291 cosmids was carried out at 65 °C overnight and washes were performed at the hybridization temperature. For the 17 cosmids that were used in the second screen, the hybridization was carried out at 65 °C, but for only 8 h, and the time of exposure to the film was decreased.

Detection of biotin-containing proteins by Western blotting

Cell extracts from *C. glutamicum* were prepared as described by Jetten and Sinskey (Jetten, M.S.M., & Sinskey, A.J., *FEMS Microbiol. Lett. 111*:183-188 (1993)). Proteins in cell extracts were separated in sodium dodecyl sulfate (SDS)/7.5% polyacrylamide gels in a BioRad mini gel apparatus and were electroblotted onto nitro-cellulose, using the BioRad mini transblot apparatus described by Towbin *et al.* (Towbin, H., *et al., Proc. Natl. Acad. Sci. USA 76*:4350-4354 (1979)). Biotinylated proteins were detected using avidin-conjugated alkaline phosphatase from BioRad and 5-bromo-4-chloro-3-indoylphosphate-p-toludine salt/nitroblue tetrazolium chloride from Schleicher & Schüll.

DNA sequencing

Automated DNA sequencing was performed by the MIT Biopolymers facility employing an ABI Prism 377 DNA sequencer.

Sequence analysis

The program DNA Strider Version 1.0 (Institut de Recherche Fondamentale, France) was used to invert, complement and translate the DNA sequence, and find open-reading frames in the sequence. The BLAST program (Altschul, S.F., *et al., J. Mol. Biol. 215*:403-410 (1990)) from the National Center for Biotechnology Information (NCBI) was employed to compare protein and DNA sequences. Homology searches in proteins were done using the MACAW software (NCBI). PCR primers were designed with the aid of the Primer Premier software from Biosoft International. The compute pI/MW tool on the ExPasy molecular biology server (University of Geneva) was used to predict the molecular mass and pI of the deduced amino acid sequence.

Example 1: Western blotting to detect biotinylated enzymes

Since pyruvate carboxylate is known to contain biotin, Western blotting was used to detect the production of biotinylated proteins by *C. glutamicum*. Two biotinylated proteins were detected in extracts prepared from cells grown in LB medium, (data not shown) consistent with previous reports. One band, located at approximately 80 kDa, has been identified as the biotin-carboxyl-carrier domain (BCCP) of the acetyl-CoA carboxylase (Jager, W., *et al., Arch. Microbiol. 166*:76-82 (1996)). The second band, at 120 kDa, is believed to be the pyruvate carboxylase enzyme, as these proteins are in the range 113-130 kDa (Attwood, P.V., *Int. J. Biochem. Cell. Biol. 27*:231-249 (1995)).

Example 2: PCR and cloning

C. glutamicum pyruvate carboxylase gene was cloned on the basis of the homology of highly conserved regions in previously cloned genes. Pyruvate carboxylase genes from thirteen organisms were examined and primers corresponding to an ATP-binding submotif conserved in pyruvate carboxylases and the region close to the pyruvate-binding motif (Table 1) were designed. Where the amino acids were different the primers were designed on the basis of *M. tuberculosis* because of its close relationship to *C. glutamicum*. An 850-bp fragment was amplified from *C. glutamicum* genomic DNA using the PCR and cloned in the pCR2.1 TOPO vector of Invitrogen to construct plasmid

pRR850. Primers were also designed based on the conserved biotin-binding site and pyruvate-binding site (data not shown).

Example 3: Isolating a cosmid containing the C. glutamicum pyruvate carboxylase gene

The 850-base-pair fragment containing a portion of the *C. glutamicum* pyruvate carboxylase gene was used to probe a *C. glutamicum* genomic library. In the first round of screening, 17 out of 291 cosmids in a dot blot appeared positive. A second round of screening was performed on these 17 cosmids, using the same probe but more stringent hybridization conditions, yielding four cosmids with a positive signal. To confirm that these cosmids indeed contained the pyruvate carboxylase gene, PCR was performed using the four positive cosmids as templates and the same primers used to make the probe. An 850-bp fragment was amplified from all four positive cosmids, designated IIIF10, IIE9, IIIG7 and IIIB7.

Organism	Conserved region A	Conserved region B
Caenorhabditis elegans	YFIEVNAR	ATFDVSM
Aedes aegypti	YFIEVNAR	ATFDVAL
Mycobacterium tuberculosis	VFIEMNPR	ATYDVAL
Bacillus stearothermophilus	YFIEVNPR	ATFDVAY
Pichia pastoris	YFIEINPR	ATFDVSM
Mus musculus	YFIEVNSR	ATFDVAM
Rattus norvegicus	YFIEVNSR	ATFDVAM
Saccharomyces cerevisiae 1	YFIEINPR	ATFDVAM
Saccharomyces cerevisiae 2	YFIEINPR	ATFDVAM
Rhizabium etli	YFIEVNPR	ATFDVSM
Homo sapiens	YFIEVNSR	ATFDVAM
Schizosaccharomyces pombe	YFIEINPR	ATFDVSM

Table 1 Pyruvate carboxylase sequences from 13 organisms (obtained from GenBank) were aligned using the MACAW software. Two highly conserved regions were selected and oligonucleotide primers were designed on the basis of the *Mycobacterium tuberculosis* DNA sequence corresponding to these regions. The forward primer was based on the DNA sequence corresponding to conserved region A and the reverse primer was based on the DNA sequence corresponding to conserved region B.

Example 4: Sequencing strategy

The 850-bp insert of plasmid pRR850 was sequenced using the M13 forward and M13 reverse primers. On the basis of this sequence, primers Begrev1 and Endfor1 were designed and used to sequence outwards from the beginning and the end of the 850-bp portion of the pyruvate carboxylase gene. Cosmid III F10 was used as the sequencing template. The sequencing was continued by designing new primers (Table 2) and "walking" across the gene.

Example 5: Sequence analysis

3637 bp of cosmid III F10 were sequenced. A 3420-bp open reading frame was identified, which is predicted to encode a protein of 1140 amino acids. The deduced protein is 63% identical to *M. tuberculosis* pyruvate carboxylase and 44% identical to human pyruvate carboxylase, and the *C. glutamicum* gene *pc* was named on the basis of this homology. The deduced protein has a predicted pI of 5.4 and molecular mass of 123.6 kDa, which is similar to the subunit molecular mass of 120 kDa estimated by SDS/polyacrylamide gel electrophoresis. Upstream of the starting methionine there appears to be a consensus ribosome binding-site AAGGAA. The predicted translational start site, based on homology to the *M. tuberculosis* sequence, is a GTG codon, as has been observed in other bacterial sequences (Stryer, L., *Biochemistry*, 3rd edn., Freeman, NY (1988); Keilhauer, C., *et al., J. Bacteriol. 175*:5595-5603 (1993)). The DNA sequence has been submitted to GenBank and has been assigned the accession number AF038548.

The amino-terminal segment of the *C. glutamicum* pyruvate carboxylase contains the hexapeptide GGGGRG, which matches the GGGG(R/K)G sequence that is found in all biotin-binding proteins and is believed to be an ATP-binding site (Fry, D.C., *et al., Proc. Natl. Acad. Sci. USA 83*:907-911 (1986); Post, L.E., *et al., J. Biol. Chem. 265*:7742-7747 (1990)). A second region that is proposed to be involved in ATP binding and is present in biotin-dependent carboxylases and carbamyphosphate synthetase (Lim, F., *et al., J. Biol. Chem. 263*:11493-11497 (1988)) is conserved in the *C. glutamicum* sequence. The predicted *C. glutamicum* pyruvate carboxylase protein also contains a

putative pyruvate-binding motif, FLFEDPWDR, which is conserved in the transcarboxylase domains of *Mycobacterium, Rhizobium* and human pyruvate carboxylases (Dunn, M.F., *et al., J. Bacteriol. 178*:5960-5970 (1996)). Tryptophan fluorescence studies with transcarboxylase have shown that the Trp residue present in this motif is involved in pyruvate binding (Kumer, G.K., *et al., Biochemistry 27*:5978-5983 (1988)). The carboxy-terminal segment of the enzyme contains a putative biotin-binding site, AMKM, which is identical to those found in other pyruvate carboxylases as well as the biotin-carboxyl-carrier protein (BCCP) domains of other biotin-dependent enzymes.

Primer name	Primer sequence (5'-3')
Begrev1	TTCACCAGGTCCACCTCG
Endfor1	CGTCGCAAAGCTGACTCC
Begrev2	GATGCTTCTGTTGCTAATTTGC
Endfor2	GGCCATTAAGGATATGGCTG
Begrev3	GCGGTGGAATGATCCCCGA
Endfor3	ACCGCACTGGGCCTTGCG
Endfor4	TCGCCGCTTCGGCAACAC

Table 2 DNA sequences of the primers used to obtain the sequence of the pyruvate carboxylase gene in the cosmid IIIF10

Previous studies have shown that phospho*enol*pyruvate carboxylase (*ppc*) is not the main anaplerotic enzyme for *C. glutamicum*, since its absence does not affect lysine production (Gubler, M., *et al., Appl. Microbiol. Biotechnol. 40*:857-863 (1994); Peters-Wendisch, P.G., *et al., Microbiol. Lett. 112*:269-274 (1993)). Moreover, a number of studies have indicated the presence of a pyruvate-carboxylating enzyme, employing ^{13}C-labeling experiments and NMR and GC-MS analysis (Park, S.M., *et al., Applied Microbiol.. Biotechnol. 47*:430-440 (1997b); Peters-Wendisch, P.G., *et al., Arch. Microbiol. 165*:387-396 (1996)), or enzymatic assays with cell free extracts (Tosaka, O., *Agric. Biol. Chem. 43*:1513-1519 (1979)) and permeable cells (Peters-Wendisch, P.G., *et al., Microbiol. 143*:1095-1103 (1997)). Very low pyruvate carboxylation activity were detected in cell-free extracts, but this activity was not uncoupled from a very high ATP background. It is highly probable that the activity measured is due to reversible gluconeogenic enzymes, such as oxaloacetate decarboxylase and malic enzyme. The presence of pyruvate carboxylase in *C. glutamicum* makes it highly unlikely that the

gluconeogenic enzymes mentioned above can serve the anaplerotic needs of this strain.

The deduced amino acid sequence of the *C. glutamicum* pyruvate carboxylase gene has significant similarity to the pyruvate carboxylase sequences from a diverse group of organisms. It contains a biotin carboxylase domain in its N-terminal region, a BCCP domain in its C-terminal region, and a transcarboxylase domain with a binding site specific for pyruvate in its central region. The *C. glutamicum* pyruvate carboxylase protein showed strong homology to *M. tuberculosis* and the human pyruvate carboxylase (Wexler, I.D., *et al., Biochim. Biophys. Acta 1227*:46-52 (1994)).

There are precedents to finding that *C. glutamicum* contains more than one enzyme to perform the anaplerotic function of regenerating oxaloacetate. *Pseudomonas citronellolis, Pseudomonas fluorscens, Azotobacter vinelandii* and *Thiobacillus novellus* contain both *ppc* and pyruvate carboxylase (O'Brien, R.W., *et al., J. Biol. Chem. 252*:1257-1263 (1977); Scrutton, M.C. and Taylor, B.L., *Arch. Biochem. Biophys. 164*:641-654 (1974); Milrad de Forchetti, S.R., & Cazullo, J.J., *J. Gen. Microbiol. 93*:75-81 (1976); Charles, A.M., & Willer, D.W., *Can. J. Microbiol. 30*:532-539 (1984)). *Zea mays* contains three isozymes of *ppc* (Toh, H., *et al., Plant Cell Environ. 17*:31-43 (1994)) and *Saccharomyces cerevisiae* contains two isozymes of pyruvate carboxylase (Brewster, N.K., *et al., Arch. Biochem. Biophys. 311*:62-71 (1994)), each differentially regulated. With the present discovery of the existence of a pyruvate carboxylase gene in *C. glutamicum*, the number of enzymes that can interconvert phospho*enol*pyruvate (PEP), oxaloacetate and pyruvate in this strain rises to six. This presence of all six enzymes in one organism has not been reported previously. *P. citronellolis* contains a set of five enzymes that interconvert oxaloacetate, PEP and pyruvate, namely pyruvate kinase, PEP synthetase, PEP carboxylase, oxaloacetate decarboxylase and pyruvate carboxylase (O'Brien, R.W., *et al., J. Biol. Chem. 252*:1257-1263 (1977)). *Azotobacter* contains all of the above enzymes except PEP synthetase (Scrutton, M.C., & Taylor, B.L., *Arch. Biochem. Biophys. 164*:641-654 (1974)).

The presence in *C. glutamicum* of the six metabolically related enzymes suggests that the regulation of these enzymes through effectors is important. Biochemical and genetic study of all six enzymes in coordination with other downstream activities may lead to the elucidation of the exact procedures necessary for maximizing the production of primary metabolites by this industrially important organism.

Example 6: Construction of a pyruvate carboxylase mutant

The entire reading frame from nucleotide 180 to nucleotide 3630 of the pyruvate carboxylase DNA was amplified using PCR. The oligonucleotide primers used for the PCR were designed to remove the *Sal*I site within the coding sequence by silent mutagenesis and introduce *Eco*RV and *Sal*I sites upstream and downstream, respectively, of the open reading frame. The PCR product was digested with *Eco*RV and *Sal*I and cloned into the vector pBluescript. The resulting plasmid is pPCBluescript. To obtain a plasmid-borne disruption of *pyc*, a derivative of pPCBluescript was constructed in which the middle portion of the *pyc* gene was deleted and replaced with the *tsr* gene, which encodes resistance to the antibiotic thiostrepton. The RP4 *mob* element was then inserted into the plasmid, yielding pAL240. This plasmid can be conjugally transferred into *Corynebacterium*, but it is then unable to replicate because it has only a ColE1 origin of replication. pAL240 was transferred from *E. coli* S17-1 into *C. glutamicum* via transconjugation, and transconjugants were selected on medium containing thiostrepton and nalidixic acid.

After the drug resistance phenotype of each transconjugant was confirmed, the transconjugants were tested for their ability to grow on different carbon sources. Because pAL240 cannot replicate in *C. glutamicum*, the only cells which will survive should be those whose genomes have undergone recombination with the plasmid. Several candidates were identified with the proper set of phenotypes: they are resistant to thiostrepton and nalidixic acid, grow well on minimal plates containing glucose or acetate as the sole carbon source, and grow poorly or not at all on minimal plates containing lactate as the sole carbon source. Southern hybridization and PCR-based assays are used to confirm whether there is only one copy of the pyruvate carboxylase gene in the genome and that it is disrupted with the thiostrepton resistance marker. Lysine production and the production of biotinylated proteins by this strain is examined, and the Δ*pyc* strain as a negative control in activity assays and as a host strain for complementation tests.

Example 7: Development of an overexpressing strain

In order to test the hypothesis that increased levels of pyruvate carboxylase will lead to increased production of lysine, it is necessary to construct strains in which expression of the pyruvate carboxylase gene is under the control of an inducible promoter.

The vector pAPE12, which has the NG2 origin of replication and a multiple cloning site downstream of the IPTG-controlled *trc* promoter, was used as an expression vector in *C. glutamicum*. A derivative of pAPE12 was constructed which contained the pyruvate carboxylase gene downstream of Ptrc. The *pyc* gene was excised from pPCBluescript using *Sal*I and *Xba*I and ligated into pAPE12 which had been cleaved with the same enzymes, forming pLW305. The pyruvate carboxylase gene present in PCBluescript (and hence in pLW305) has the wild type GTG start codon, and the *Sal*I restriction site present near the 5' end of the wild type gene was eliminated by the introduction of a one base silent mutation during amplification of the pyruvate carboxylase gene.

pLW305 and pAPE12 was electroporated into several other *Corynebacterium* genetic backgrounds.

Because the pyruvate carboxylase gene in pLW305 has a GTG start codon and carries some intervening DNA between the trc promoter and the start codon, a pyruvate carboxylase overexpression plasmid, pXL1, was designed that eliminates those shortcomings. The 5' end of the gene was amplified from pLW305 with oligonucleotide primers that simultaneously change the GTG start codon to ATG and introduce a *Bsp*LU11-I restriction site, which is compatible with *Nco*I. The PCR product was then cut with *Bsp*LU11-I and *Afe*I, and ligated into the 7.5 kb backbone obtained by partial digest of pLW305 with *Nco*I followed by complete cutting with *Afe*I. Two independent sets of ligations and transformations have yielded putative pXL1 clones.

Example 8: Fermentation results

It has been shown that the level of pyruvate carboxylase activity varies greatly with the carbon source used when the gene is expressed from its native C.

glutamicum promoter. Therefore, production of pyruvate carboxylase in strains grown on these carbon sources was examined.

The strains NRRL B-11474, NRRL B-11474 (pLW305), and NRRL B-11474 Δ*pyc* candidate 35 were cultured in flasks on minimal medium for NRRL B-11474 with two different sources of carbon: glucose or lactate. The results on growth and amino acid production are presented below.

	glucose			lactate		
	biomass (g/l)	lysine (g/l)	Y lys/glc (g/g)	biomass (g/l)	lysine (g/l)	Y lys/lac (g/g)
NRRL B-11474	6.7 ± 0.2	5.0 ± 0.7	0.21	3	1.7	0.12
NRRL B-11474 (pLW305)	7.3 ± 0.2	5.3 ± 0.2	0.22	4	2.5	0.15
Δ*pyc* #35	1.1	0	0	0	0	0

NRRL B-11474 and pLW305 show the same behavior on glucose. Both strains produce the same amount of biomass and lysine. On lactate the strains also have similar yield of lysine. NRRL B-11474 (pLW305) consumed all of the lactate in the medium (17g/l) whereas the wild type NRRL B-11474 consumed 40% less lactate during the same period of time. The NRRL B-11474 was calculated to consume lactate at a rate of 0.37 g lactate/hour, whereas the NRRL B-11474 (pLW305) strain consumed this substrate at a rate of 0.65 g lactate/hour.

The NRRL B-11474 Δ*pyc* doesn't grow on lactate, which is consistent with the expected phenotype. Its growth on glucose is very low and the strain does not produce lysine. Kinetic studies are conducted to characterize further the behavior of these strains.

Example 9: Visualization of biotinylated proteins

Pyruvate carboxylase contains biotin. Therefore, it should be possible to detect the accumulation of this enzyme by monitoring the appearance of specific

biotinylated products in cells.

Example 10: Electrophoretic gels

To detect biotinylated proteins in electrophoretic gels, a commercially available streptavidin linked to alkaline phosphatase was used. Crude protein lysates from induced and uninduced cultures of *E. coli* DH5α or NRRL B-11474 harboring pAPE12 or pLW305 and separated the proteins on duplicate 7.5% polyacrylamide denaturing electrophoretic gels. One gel of each pair is stained with Coomassie Brilliant Blue to visualize all proteins and ensure equal levels of protein were loaded in each lane. The other gels are treated with the streptavidin-alkaline phosphatase reagent, which binds to biotinylated proteins. The location of these proteins can then be visualized by providing alkaline phosphatase with a colorimetric substrate, 5-bromo-4-chloro-3-indolyl phosphate (BCIP). As reported by others, two major biotinylated proteins were detected. The higher molecular weight species (approx. 120 kDa) has been shown to be pyruvate carboxylase, and the lower molecular weight species (approx. 60 kDa) is the biotinylated subunit of acetyl-CoA carboxylase.

All publications mentioned hereinabove are hereby incorporated in their entirety by reference.

While the foregoing invention has been described in some detail for purposes of clarity and understanding, it well be appreciated by one skilled in the art from a reading of this disclosure that various changes in form and detail can be made without departing from the true scope of the invention and appended claims.